E42

(19) **JAPANESE PATENT OFFICE**

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **10320409 A**

(43) Date of publication of application: **04.12.98**

(51) Int. Cl	G06F 17/30
	G06F 17/27

(21) Application number: 09128986	(71) Applicant: SEIKO EPSON CORP
(22) Date of filing: 19.05.97	(72) Inventor: MIWA SHINJI

(54) **METHOD AND DEVICE FOR EXTRACTING DOCUMENT INFORMATION AND STORAGE MEDIUM STORING DOCUMENT EXTRACTING PROCESS PROGRAM**

(57) Abstract

PROBLEM TO BE SOLVED: To extract a group of document contents as a context since it is hard to extract part of a document or find differences between two documents, if a document is fractionized too much.

SOLUTION: Paragraphs are detected from the document contents of one document, the document contents are divided by the paragraphs, and a morpheme analysis is carried out by the paragraphs. Then featured elements are extracted on the basis of the morpheme analytic results (step s1) and a feature table is generated which shows the relation ship between the featured elements and the paragraphs including the featured elements (step s2). On the basis of this feature table, the document is classified by contents as meaningful groups (step s3) and when a content selection indication is received from a user (step s4), the document contents of the paragraph belonging to the selected content are outputted (steps s5 and s6).

COPYRIGHT: (C)1998,JPO

-1-